2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT) | 978-1-6654-5361-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/TQCEBT54229.2022.10041629

Classification of Disaster Tweets using Machine Learning and Deep Learning Techniques

Asinthara K
Christ University
Bangalore
asintharak@gmail.com

Meghna Jayan
Christ University
Bangalore
meghnajayan2801@gmail.com

Lija Jacob
Christ University
Bangalore
lija.jacob@christuniversity.in

***Abstract*—Social networks provide a plethora of information for gathering extra data on people's behavior, trends, opinions, and feelings during human-affecting occurrences, such as natural catastrophes. Twitter is an inevitable communication medium during calamities. People mainly depend on Twitter to announce real-time emergencies. However, it is rarely straightforward if someone is declaring a tragedy. Sentiment analysis of disaster tweets aid in situational awareness and realizing the disaster dynamics. In our paper, we perform a sentimental analysis of disaster tweets using techniques based on machine learning and deep learning. The tweets are pre-processed before being converted into a structured form using Natural Language Processing (NLP) methods. Supervised learning techniques such as the Support Vector Machine and the Naïve Bayes Classifier algorithm are used to develop the Classifier, which categorizes tweets into distinct catastrophes and selects the most appropriate algorithm. The chosen algorithm is further enriched with an emoticon detection algorithm for explicit elucidation. Our research would help disaster relief organizations and news agencies to conclude about the state of affairs and do the needful.**

Keywords—Twitter, disaster tweets, disaster management, natural language processing, machine learning, deep learning

I. Introduction

Microblogging is a common form of communication among internet user's today. People frequently switch to microblogging services because of the free nature of messages and the simplicity of access to microblogging platforms. As more people tweet about the products and services used as well as the political and religious perspectives, microblogging websites are becoming key contributors to public sentiment. There are numerous well-known microblogging websites, such as Twitter, Facebook, Instagram, and others. Microblogging helps us to communicate short content more quickly and helps businesses to disseminate information that needs to be shared quickly in a fast-paced setting.

Microblogs are used to communicate information and monitor broad public opinion during human-affecting events, such as sporting events, political elections, natural disasters, etc., [2]. As a microblogging website, Twitter has become a crucial communication tool during emergencies. The small messages (called "tweets") sometimes reveal users' sentiments during different situations. In recent years, especially during a natural disaster that typically happens with no warning, Twitter has gained a lot of attention as a brand-new platform for crisis communication where users are live-broadcasting news, images, and opinions from the crisis area. [2].

Natural Language Processing (NLP), a subfield of artificial intelligence, enables computers to comprehend and analyze text. Computational linguistics, along with statistical and machine learning methods, enables a system to understand the meaning of a text. The developing method of sentiment analysis uses NLP to determine how a text makes you feel. The sentiment analysis of a tweet related to a disaster can help the world realize the worse condition of the place or the problem and allow emergency management agencies to start the rescue operation profoundly. Emergency responders can better comprehend the system's dynamics by gathering and analyzing social media information and mining users' spatial sentiments and opinions during a crisis. One can find out, for instance, about the primary users' anxieties and panics, the personal effects of user connections, and the geographical locations most significantly impacted by the crisis [7].

A lot of research has happened in Sentiment analysis of disaster tweets. Most of them focused on identifying themes and classifying sentiments based on the similarity and polarity of tweets. The majority concentrated on finding themes and categorizing attitudes based on how similar and diametrically opposed tweets were. To recognize, categorize, and predict disaster-related tweets, deep learning and supervised machine learning approaches are applied. Most studies employ machine learning techniques including Naïve Bayes Classifier, Support Vector Machine, Maximum Entropy, and deep learning technique LSTM. Our proposed methodology is to build a sentiment classifier to analyze the tweets and classify them into various disasters using machine learning models like Naïve Bayes Classifier and Support Vector Machine and to analyze the severity of the disaster. A comparison analysis is conducted between these classifiers to choose the best among them and add a further algorithm to identify emoticons used in the tweets, giving valuable insights.

II. Literature Review

Kevin Stove et al. [1] tested models such as Support vector machines, Maximum entropy, and Naïve Bayes classifier on catastrophic tweets in 2016 and determined that SVM gave the greatest F1 Score out of these three models. Researchers also found that it was feasible to accurately and reliably identify relevant tweets. However, the data sparsity caused difficulty in the model, and also, the Retweets and URLs were not generally valuable, likely because lexical features already captured the information. Himanshu Shekhar and Shankar Setty [2] proposed a model to extract people's emotions during and after a disaster through user Twitter posts. The contributions of the paper are to demonstrate the geographical distribution of selected natural disasters within a period that has only been taken from

Twitter, the continent-wise frequency of the disaster occurrence, and the analysis of people's sentiments during a disaster by sentiment analysis. In 2017, Si Si Mar Win and Than Nwe Aung [3] published a tweet monitoring system that uses the LibLinear classifier to categorise emergency messages into three tiers. Only disaster lexicon-based features, sentiment lexicon-based features, and linguistic features were employed in the feature extraction. Four models were implemented to determine the accuracy: Naïve Bayes, SVM, SMO, and Random Forest.Due to the sensitivity of the models, Information Gain was used for feature selection. Here the disaster was again classified as bushfire, Yolanda, Volcano, and Earthquake. The extraction technique beats both the neural word embedding model and the common bag of words model. The drawback of the model was that automated mechanisms that are relevant to finding actionable information from Twitter are not used in this paper.

TABLE I. ACCURACIES OF MODELS USED IN [3]

	LibLinear Accuracy	SMO Accuracy	Random Forest Accuracy
Bushfire in Australia	87.28%	88.28%	85.45%
Yolanda in Typhoon	92.71%	93.31%	92.32%
Volcano in Iceland	92.71%	93.31%	92.32%
Earthquake in Nepal	76.90%	76.43%	72.13%

In 2020 Jayashree Domala et al. [4] created a software solution program that enables crisis management sites to dynamically display news about disasters, distributed to other social media accounts via their websites. The architecture is based on machine learning approaches and classifies news from prominent media websites into two categories: disaster-related news and non-disaster-related news. The implementation of the algorithm using a bag of words representation is considered both unigram and bigram. Several classification approaches, including logistic regression, multinomial Naïve Bayes, Xtreme gradient boosting, SVM, and random forest, were used. The best outcome produced was the logistic regression model. In 2021 Aryan Karnati et al. [5] utilized the Bidirectional long-term memory(LSTM) method to classify tweets.LSTM was used to run inputs in two ways which help preserve information from the past and future. The model produced an accuracy of 85.6%, which is comparatively higher than traditional machine learning algorithms.

Using machine learning and natural language processing techniques, Sudha Verma et al. [6] created a classifier in 2011 that distinguishes tweets based on impersonal or personal style, subjectivity, and linguistic register. The authors demonstrate how a low-level linguistic variables classifier can successfully identify tweets that add to situational awareness. Additionally, how language motivations improve the performance of the system was also demonstrated. These findings show that important user behavior characteristics can forecast whether a tweet will contain tactical information. In 2017, when the horrific Hurricane Sandy hit, Venkata K Neppalli et al. [7] did a sentiment analysis of the tweets that were sent out. Researchers then used a map with the hurricane as the center to show how the internet's sentiments were expressed. The researchers demonstrated how users' feelings might vary based on disasters far away and their physical locales. Researchers also examined how the diversity in viewpoints in a tweet sent out during the disaster affects its likelihood to be retweeted. The feature extraction was done using Unigrams, Polarity clues, Emoticons, and SentiStrength. The performance of the classifiers Naive Bayes and Support Vector Machine is described in table II.

TABLE II. CLASSIFIERS USED AND THEIR ACCURACIES [3]

Feature type	**Naive Bayes**	**SVM**			
		C=0.1	**C=0.5**	**C=0.75**	**C=1**
Sentiment-based	68.6	67.95	67.52	67.09	67.09
Unigrams	71.82	72.25	72.04	70.1	68.6
Combination	73.33	**75.91**	73.54	72.47	71.61

After finding the best model, geo-mapped sentiment analysis was used to visualize the user's performance. The users' perceptions vary by location and from disaster to disaster. And the scope of retweetability decreases as the emotional divergence increases.

In 2010, Alexander Pal and Patrick Paroubek [8] presented a method of automatic collection of the corpus used for sentiment analysis. TreeTagger for POS-Tagging was able to differentiate the polarity of tweets. The classifier uses the multinomial Naive Bayes classifier, which accepts features like POS tags and N-grams.. The two strategies for filtering out the common n-grams were: salience and entropy, and using salience got better accuracy. In 2018 Kevin Stowe et al. [9] proposed a model that focuses on Twitter users that are potentially at risk but are difficult to identify due to the noisiness of Twitter data. Three clustering methods were used to reduce noisiness: spatial clustering, Spatio-Temporal clustering, and Temporal clustering. Here the tweets are displayed in chronological order, colored by different clusters. For classifying, the ensemble method with 20 classifiers was used and was able to get an F1 score of 83% over the performance of a single classifier with 69%. The two models mainly used were multi-layer perceptron and Convolutional neural network.

In 2021 Berenice Jacqueline Sanchez Alvarado and Pedro Esteban Chavarrias Solano [10] presented a natural language approach for solving the task of detecting disaster tweets. The problem of sentiment analysis has been approached in a variety of ways including supervised learning algorithms, such as Maximum Entropy, Naïve Bayes, and Support vector machines. Among the other unsupervised learning methods like lexicons, grammatical analysis, and syntactical analysis. The methodology used in the model is summarized in fig. 1. using chatbots. For NLP

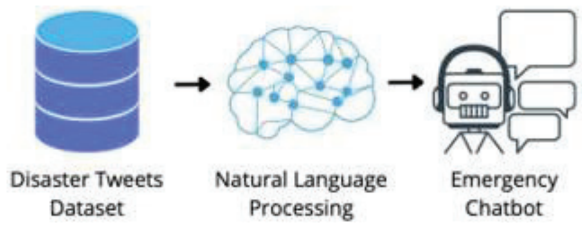

Fig. 1. Emergency chatbot [10]

RoBERTa (Robustly optimized BERT) approach was used and achieved an accuracy of 0.97%.

In 2017, Ahmed Nagy and Jeannie Stamberger [11] created techniques that can gauge the mood of Tweets in a disaster-prone environment and compared four different methods, which included different combinations:

1) SentiWordNet, Emoticons, and AFNN, 2) SentiWordNet and Emoticons, 3) Bayesian Network, 4) SentiWordNet, Emoticons, AFNN, AND Bayesian Network. The result produced by all four has been described in table III. The above led us to the conclusion that SentiWordNet and Bayesian Networks provided the best accuracy and recall. However, some tweets were misclassified as the dataset was small. Nasser Assery et al. [12] created a model in 2019 and employed a variety of machine learning techniques to analyze the Twitter data produced during Hurricanes Florence and Michael. Nasser Assery et al. [12] created a model in 2019 and employed a variety of machine learning techniques to analyze the Twitter data produced during Hurricanes Florence and Michael. In order to create lexical attributes for text classification, a number of machine learning techniques utilizing two vectorizers, CountVectorizer and TfidfVectorizer, were examined. In 2015 Davide Buscaldi and Irazu Hernandez-Farias [13] used the IRADABE model, which performs Sentiment and polarity classification in support vector machines, and the dataset was POS-Tagged using TreeTagger. For feature extraction, Bag-of- words, Emoticons frequency, URL information frequency, Subjectivity features, tweet length, and upper case ratio, SentiWordNet, Hu-Liu Lexicon, AFINN Lexicon, Whissel Dictionary, and Counter-Factuality Compression were used. The authors concluded that IRADABE using SVM, got an accuracy of 70% and also could identify subjective posts with good accuracy, but the insights on the specific situation will be missed when focusing on broader data and was unable to consider different feelings.

TABLE III. RESULTS PRODUCED BY THE MODELS [11]

	AFNN+ Emoticons+ SentiWordNet	**Emoticons + SentiWordNet**	**Bayesian Network**	**AFNN+ Emoticons + SentiWordNet +Bayesian Network**
Recall	0.95	0.93	0.72	0.96
Precision	0.93	0.90	0.85	0.94
F-measure	0.939	0.914	0.779	0.949

In 2014 Muhammed Imran et al. [14] developed an AIDR platform to classify crisis-related microblogs automatically. The platform uses machine and human intelligence to classify additional posts by training an artificial classifier with labels from various communications. The platform uses an active learning strategy to improve classification accuracy when additional training examples are available. The AIDR model gained an accuracy of approximately 80%. Different supervised learning techniques were used by Shivam Behl et al. [15] for the multi-class categorization of Twitter data. Using Twitter, the significance and difficulty of using social networking sites in relief work were explored. To investigate the model's reusability, the model employed two sets of data (Nepal and Italy Earthquake) during training and one original COVID-19 dataset during testing. For feature extraction, Tfidfvectorizer and Word2Vec were used. In addition, researchers contrasted cutting-edge classification systems for separating tweets about resource availability from tweets about resource demands. On COVID-19 data, MLP outperformed Word2Vec with an accuracy of 83%. To analyze the behavior of the model and to explain it, LIME was employed.

Vedurumudi Priyanka [16] attempted to apply sophisticated text mining algorithms to examine the moods of the text and categorize it as neutral, negative, and positive. The technique is also known as Opinion mining. Feature extraction uses unigram and bigram techniques to classify machine learning methods like Multi-Layer Perceptron, Naïve Bayes Classifier, Decision Tree, Maximum Entropy, Support Vector Machine, Random Forest, Convolutional neural network, XGBoost, and Recurrent neural network. Sparse representation of feature type was done by dividing it into presence and frequency tables. In the presence feature type, the vector has 1 as the index, and in the frequency feature type, 0 is the index. After applying the machine learning techniques, an ensemble learning method is applied to the model. The flowchart of the ensemble is shown in fig. 2.

Support Vector Machine (SVM) and Naive Bayes Classifier were employed by Parilla-Ferrer et al. [17] to automatically categorize tweets. The models' precision, accuracy, F1-score, recall, and area under the curve were evaluated. The collection of tweets was compiled during the 2012 Habagat Flooding in Metro Manila, and it revealed that users broadcasted more tweets expressing their thoughts and opinions since ambiguous tweets predominated over relevant ones. Throughout the experiment, stop words were eliminated using Porter's method, and the bag-of-words technique was utilized to create the key characteristics of the word vector. SVM topped Naïve Bayes with 10-fold cross-validation in the areas of accuracy, F1-score, AUC, and recall, but Naïve Bayes exceeded SVM in terms of precision. Additionally, the researchers came to the conclusion that there were 65% more uninformative tweets than informative tweets.

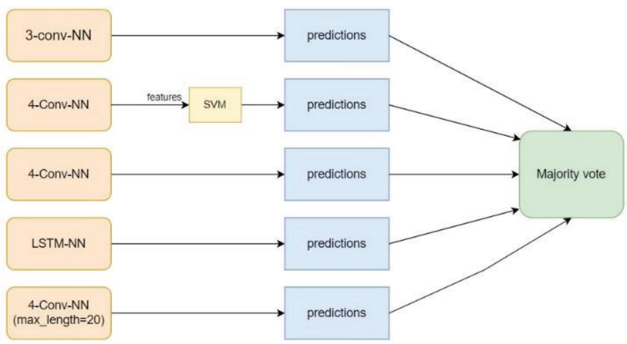

Fig. 2. Different combinations of ensemble learning methods [16]

TABLE IV. ACCURACIES OF THE MODEL [16]

Model	Accuracy
LSTM – NN	83.00
4-Conv-NN	83.34
4-Conv-NN features + SVM	83.39
4-Conv-NN with max_length - 20	82.85
3-Conv-NN	82.95
Majority Vote Ensemble	83.58

Agarwal et al. [18] aimed to build models for two different tasks. Tweets are categorized in two ways: binary (positive or negative) and three-way (positive, negative, or neutral). The experiment was done with the unigram model, a feature-based
model, and a tree kernel model, as well as two combinations of these three model typesA vocabulary of acronyms and emoticons was also introduced in the study. The base model (unigram) was outperformed by the feature-based and the tree kernel model. Based on the analysis conducted for a feature-based model, the authors concluded that the most significant features integrate the previous polarity of words and their parts of speech tags. However, Go et al.[19] classified the tweets positively or negatively based on a query term using machine learning algorithms like Naïve Bayes Classifier, Support Vector Machine, and Maximum Entropy to do the same and obtained more than 80% accuracy.

Zahra Ashktorab et al. [20] unveiled a Twitter mining method Tweedr in 2014. Tweedr gathers data regarding natural calamities from organizations that provide aid. The three components of the Tweedr pipeline are classification, clustering, and extraction. SVM, sLDA, and Logistic regression are just a few of the classification techniques used in the classification phase. Some filters were employed in the clustering phase to combine similar tweets, then, during the extraction stage, tokens and words that specifically mention various categories of damage types, infrastructure damage, and casualties were extracted. In the classification phase, Logistic Regression got an accuracy of 0.86%, the highest among all the models. In the extraction phase, conditional random fields implemented by the CRFsuite toolkit were used, and in the clustering phase, two different methods of approximate string matching: Bloom filters and SimHash were considered.

TABLE V. EXAMPLE OF ACRONYM DICTIONARY USED IN [18]

Acronym	English Expansion
gr8, gr8t	great
lol	Laughing out loud
rotf	Rolling on the floor
bff	Best friend forever

III. METHODOLOGY

The rising accessibility of digital information is primarily responsible for the increasing popularity of NLP. The proposed methodology for the research study uses machine learning methods, such as Naïve Bayes Classifier and Support Vector Machine for multi-class classification of disasters. We also propose an emoji classification for a deeper understanding of the situation.

The primary features of the proposed method are described in the following section.

A. Collection of Tweets:

The Twitter Streaming API was mainly used to extract data. The API can download Twitter updates from the internet and convert them into String objects [2]. Using appropriate keywords or hashtags is the most excellent way to find the most significant tweets during a catastrophe or crisis [3]. Some of the relevant tweets for earthquakes are given in Table VI.

B. Preprocessing of Tweets

The process gets rid of tweets that are full of noise and repetition [3]. The goal of the cleaning process is to remove any unwanted elements from the data since these contents might lead to noise and redundancy [12].

TABLE VI. RELEVANT TWEETS FOR EARTHQUAKES [2].

Tweet	**Keywords Matched**	**Tweet Rele-vance**
wasnt in the middle school for 5 minutes today when there was an earthquake drill and i had to sit under a table for 10 minutes kewl	Earthquake, under	42 %
@geostuff - Earthquake rattles Midlothian - Edinburgh Evening News: Edinburgh Evening NewsEarthquake rattles MidlothianEdi... http://t.co/olRFvQFF3J	Earthquake, rattles, news	98 %
@ToranNigrelli - Today is the day people this is not a drill! 3pt Shootout in the Richter at 3. Regis-tration at 2:30, 10 for a two person team...BE THERE!!	Richter	12 %

C. Proposed Model

1) Support Vector Machine: An effective supervised machine learning technique that employs minimal computing power. The classification technique detects a hyperplane in an N-dimensional space that categorizes the data points with clarity. The hyperplane is selected so that the distance between the data points of each class is maximum.

2) Naïve Bayes Classifier: The Naive Bayes algorithm is a collection of classification methods based on the Bayes theorem. The algorithm is a family of algorithms that work on the same basic idea. The classifier works by constructing a feature vector from the words that appear in a single text and how frequently those words appear collectively in the data.

3) LSTM: Recurrent neural networks (RNN) able to learn persistent dependencies are called LSTMs (Long Short Term Memory). In order to find the underlying correlations in the given sequential data, an LSTM deals with algorithms that attempt to emulate the way a person's brain functions. The RNN computes the following hidden state for each element in the sequence, or each word in the tweet, using the most recent word embedding and prior hidden state.

A comparison will be carried out among these algorithms, and the best accurate model will be enhanced using an emoticon detection algorithm. The use of emoticons can aid in our understanding of both the precise crisis scenario and the local public's emotions.

IV. CONCLUSION

Disaster management is a crucial field because, without it, rescue and relief efforts during a disaster cannot happen quickly, worsening the harm done. The review examines the analysis and classification of disasters using data from Twitter. The majority of the publications in the review focused on using the Naïve Bayes Classifier and Support Vector Machine models to identify tweets as disaster-related or not. Disaster intensity can be determined by analyzing the tweets' sentiments and calculating how far each one is from the disaster's epicenter [7]. Emoticons were removed from the tweet because of the noisy behavior in the dataset. However, occasionally, these emoticons can also convey important information [20]. Thus in our study, we categorize disasters by several classes and use emoticons during the training of the algorithm to understand the type of disaster occurring.

The emoticons from the data are removed in most models which sometimes can give valuable insights into the exact mood of people. So, in our future work, we will build a model using Naïve Bayes Classifier, Support Vector Machine, and LSTM for multi-class disaster tweet classification. The best among these models will be chosen and enriched with an emoticon detection algorithm to get a clear idea about the scenario.

REFERENCES

[1] K. Stowe, M. J. Paul, M. Palmer, L. Palen, and K. Anderson, “Identifying and Categorizing Disaster-Related Tweets.” *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, pp. 1-6, Nov. 2016, doi: 10.18653/v1/w16-6201.

[2] H. Shekhar and S. Setty, “Disaster analysis through tweets.” *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, doi: 10.1109/icacci.2015.7275861.

[3] S. S. M. Win and T. N. Aung, “Target oriented tweets monitoring system during natural disasters.” *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Jun. 2017, doi: 10.1109/icis.2017.7959984.

[4] J. Domala, “Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing.” *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Aug. 2020, doi: 10.1109/icesc48915.2020.9156031.

[5] A. Karnati, S. R. Boyapally, and D. S. K.P, “Natural Language Processing with Disaster tweets using Bidirectional LSTM.” [Online]. Available: https://www.xajzkjdx.cn/gallery/1-may2021.pdf.

[6] S. Verma, “Natural Language Processing to the Rescue? Extracting ‘Situational Awareness’ Tweets During Mass Emergency.” *Proceedings of the International AAAI Conference on Web and Social Media*.

[7] V. K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, and S. Stehle, “Sentiment analysis during Hurricane Sandy in emergency response.” *International Journal of Disaster Risk Reduction*, vol. 21, pp. 213-222, Mar. 2017, doi: 10.1016/j.ijdrr.2016.12.011.

[8] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining.” *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, May 2010.

[9] K. Stowe, J. Anderson, M. Palmer, L. Palen, and K. Anderson, “Improving Classification of Twitter Behavior During Hurricane Events.” *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pp. 67-75, Jul. 2018, doi: 10.18653/v1/w18-3512.

[10] B. J. S. Alvarado and P. E. C. Solano, “Detecting Disaster Tweets using a Natural Language Processing technique.” Nov. 2021.

[11] A. Nagy and J. Stamberger, “Crowd Sentiment Detection during Disasters and Crises.” *Proceedings of the 9th International ISCRAM Conference* , Apr. 2012.

[12] N. Assery, Y. Xiaohong, S. Almalki, R. Kaushik, and Q. Xiuli, “Comparing Learning-Based Methods for Identifying Disaster- Related Tweets.” *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, doi: 10.1109/icmla.2019.00295.

[13] D. Buscaldi and I. Hernandez-Farias, “Sentiment Analysis on Microblogs for Natural Disasters Management.” *Proceedings of the 24th International Conference on World Wide Web*, pp. 1185-1188, May 2015, doi: 10.1145/2740908.2741727.

[14] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, “AIDR: Artificial Intelligence for Disaster Response.” *Proceedings of the 23rd International Conference on World Wide Web*, pp. 159-162, Apr. 2014, doi: 10.1145/2567948.2577034.

[15] S. Behl, A. Rao, S. Aggarwal, S. Chadha, and H. Pannu, “Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises.” *International Journal of Disaster Risk Reduction*, vol. 55, p. 102101, Mar. 2021, doi: 10.1016/j.ijdrr.2021.102101.

[16] V. Priyanka, “Twitter Sentiment Analysis.” *i-manager’s Journal onComputer Science*, 2021.

[17] B. E. Parilla-Ferrer, P. Fernandez, and J. T. Ballena IV, “Automatic Classification of Disaster-Related Tweets.” *International conference on Innovative Engineering Technologies (ICIET’2014)*, Dec. 2014, doi: 10.15242/iie.e1214072.

[18] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment Analysis of Twitter Data.” *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pp. 30-38, Jun. 2011.

[19] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification using Distant Supervision.” 2009.

[20] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta, “Tweedr: Mining Twitter to Inform Disaster Response.” *Proceedings of the 11th International ISCRAM Conference*, May 2014.